

МЕТОД БІНАРНИХ ДЕСКРИПТОРІВ ДЛЯ ТРЕКІНГУ ОБЛИЧ

О. С. Павлусь^{1, а}, В. М. Степаненко¹

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
Фізико-технічний інститут

Анотація

У даній роботі розглянуто задачу трекінгу (відстеження) безлічі облич в потоці. Розроблено систему трекінгу, яка орієнтована на вирішення завдань автоматизованого відеонагляду. При цьому специфіка умов роботи камер відеоспостереження використана для підвищення ефективності системи в порівнянні з аналогами загального призначення. Алгоритм трекінгу заснований на бінарних дескрипторах BRIEF, які вкрай ефективно обчислюються на сучасних процесорних архітектурах.

Ключові слова: трекінг облич, бінарні дескриптори, відеонагляд

Вступ

Трекінг (відстеження) об'єктів – це складне і практично значуще завдання комп'ютерного зору. Розробка компактних, потужних і відносно недорогих пристроїв запису і обробки відеоінформації сприяє постійному розширенню сфери застосування систем трекінгу, яка включає відеоспостереження, різні мобільні додатки, доповнену реальність, спостереження за дорожнім рухом, аудіовізуальні додатки, робототехніку. Така різноманітність можливих областей застосування означає, що трекінг об'єктів являє собою цілий клас задач з різними вимогами до конкретної системи. Таким чином, існування універсального методу трекінгу малоімовірно.

Технології розпізнавання облич застосовуються в найрізноманітніших сферах:

- забезпечення безпеки в місцях великого скупчення людей;
- системи охорони, уникнути незаконного проникнення на територію об'єкта, пошук зловмисників;
- фейс-контроль в сегменті громадського харчування та розваг, пошук підозрілих і потенційно небезпечних відвідувачів;
- верифікація банківських карт;
- онлайн-платежі;
- контекстна реклама, цифровий маркетинг, Intelligent Signage і Digital Signage;
- фототехніка;
- криміналістика;
- телеконференції;
- мобільні додатки;
- пошук фото у великих базах фотознімків;
- відмітка людей на фото в соціальних мережах і багато інших.

Незважаючи на велику кількість наукових праць, опублікованих за темою трекінгу обличчя за останні

тридцять років, в даний час проблему відстеження обличчя не можна назвати вирішеною в обсязі, що дозволяє здійснювати вибір методів, алгоритмів і програмних засобів трекінгу обличчя людини на відеокадрах в інтересах різноманітних застосувань і, зокрема, для людино-машинних інтерфейсів, які розширюють інтелектуальні можливості комп'ютерів і комп'ютерних систем. Це обумовлено багатьма факторами, що ускладнюють аналіз зображення обличчя людини на відеокадрах, такими як нерівномірність освітлення сцени, непередбачуваність траєкторії і швидкості руху голови, нестабільність форми і текстури голови за рахунок зміни міміки обличчя тощо [1].

1. Аналіз існуючих методів алгоритмів трекінгу

На сьогоднішній день розроблено значну кількість алгоритмів трекінгу.

Велика частина існуючих методів трекінгу осіб заснована або на побудові моделі зовнішнього вигляду відслідковується об'єкта або фону, або на прогнозі траєкторії руху.

Більшість алгоритмів трекінгу припускають побудову опису зовнішнього вигляду об'єкта, яке використовується для оцінки його положення. Залежно від конкретної області застосування, це опис може бути статичним або динамічним.

Деякі підходи також застосовують опис фону для поліпшення якості трекінгу. Як правило, в такому випадку для відділення відслідковується об'єкта від фону використовується бінарний класифікатор. Такий підхід лежить в основі алгоритмів Ensemble Tracking і Tracking-Learning-Detection (TLD).

Ряд методів трекінгу включає в себе модель руху об'єкта. Даний підхід представлений алгоритмами на основі фільтра Калмана і фільтрів частинок (Particle Filters).

^аalex.pavlus23@gmail.com

Багато існуючих систем трекінгу в явній або неявній формі використовують комбінацію перерахованих підходів. Наприклад, алгоритми на базі широко відомого методу Лукаса-Канаді (Lukas-Kanade) включають як модель зовнішнього вигляду об'єкта, так і модель руху.

2. Постановка задачі

Представлена система трекінгу безлічі облич, призначена для вирішення завдань відеоспостереження. Існує кілька труднощів, характерних для даної сфери застосування. Перш за все, має місце згадана вище проблема висковзання треку. Крім того, в умовах сценарію трекінгу безлічі облич можливі помилки невідповідності, що виражаються в тому, що трекер плутає обличчя з пересічними траєкторіями. І нарешті, постійна необхідність детектування нових облич, що потрапляють в поле зору камери, може спричинити суттєве погіршення швидкодії системи трекінгу.

З іншого боку, специфічні особливості умов роботи камер відеоспостереження можуть бути використані для підвищення ефективності систем трекінгу. Наприклад, має сенс вважати обличчя жорсткими об'єктами і для побудови простору ознак опису особи використовувати прості алгоритми, які не потребують значних обчислювальних витрат.

Представлена система складається з детектора облич і підсистеми трекінгу і відрізняється простою і ефективністю. Алгоритм трекінгу заснований на бінарних дескрипторах. На сьогоднішній день запропоновано кілька підходів, заснованих на дескрипторах. Розглянемо дескриптори BRIEF (Binary Robust Independent Elementary Features), основна особливість яких полягає у високій обчислювальній ефективності [2].

Метод трекінгу з використанням BRIEF, заснований на комбінації статичних і динамічних словників. Підхід, запропонований в даній роботі, не має на увазі застосування подібних словників і використовує BRIEF виключно для оцінки міжкадрового переміщення облич.

У загальному випадку задача трекінгу безлічі об'єктів ставиться таким чином. Для кожного кадру вхідного відеопотоку (як правило, в реальному часі) для кожного відслідковуваного об'єкту (особи, автомобіля, пішохода і т.д.) вказується його положення (прямокутник, еліпс, центр ваги ...), ідентифікатор і, опціонально, рівень довіри. Залежно від конкретного завдання присвоєння нового треку об'єкту, який на час зник з поля зору, може вважатися як помилкою, так і нормальною поведінкою.

В рамках запропонованого підходу такі об'єкти завжди отримують нові ідентифікатори (тобто нові треки). У разі необхідності для об'єднання треків можуть застосовуватися прості методи ідентифікації по обличчю.

3. Структура системи

Запропонована система складається з двох основних підсистем: детектора і трекера (рис. 1). Детектор обробляє відносно малу частину кадрів відеопотоку, на яких створюються нові треки і коригуються існуючі. Решта кадри обробляються трекером з використанням алгоритму, що не потребує значних обчислювальних витрат.

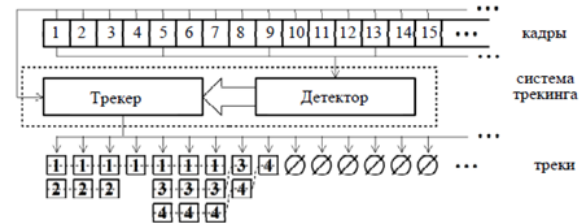


Рис. 1. Структура системи трекінгу

Термін «трекер» позначає підсистему, що відповідає виключно за підтримку треків, в той час як термін «система трекінгу» відповідає комбінації детектора і трекера.

Основною перевагою представленої на рис. 1 структури є гнучкість, так як трекер не залежить від алгоритму детектування. Крім того, можливість вибору частки кадрів, оброблюваних детектором, дозволяє досягти компромісу між швидкістю і якістю роботи. При цьому детектор може бути використаний для корекції треків, що істотно пом'якшує проблему їх ускользання. У представлений системі використовується детектор осіб, заснований на каскадних класифікаторах. У загальному випадку може бути використаний будь-який детектор.

- 1) створення та оновлення треків;
- 2) видалення треків;
- 3) міжкадрова підтримка треків.

Створення та оновлення треків. Перш за все, трекер оцінює нові положення вже облич, які відслідковуються і видаляє застарілі треки. Далі виявлені детектором обличчя використовуються для створення нових треків і оновлення існуючих. Положення виявлених облич порівнюються з наявними треками, в результаті чого складається зважений двочастковий граф (рис. 2), причому ребра з'єднують лише ті пари обличчя-трек, відносна площа перетину яких перевищує деякий заздалегідь заданий поріг. Потім будується паросочетание з максимальною кількістю ребер і мінімальною сумарною вартістю (рис. 2). Пошук такого паросполучення зводиться до вирішення задачі на знаходження потоку мінімальної вартості (minimum cost flow). У більшості випадків такий підхід дозволяє коректно вирішувати неоднозначні відповідності треків і виявлених осіб (наприклад, коли траєкторії двох обличчя перетинаються). Положення треків, яким відповідають виявлені обличчя, коригуються. Обличчя, яким не відповідає жоден трек, використовуються для створення нових треків. Треки, яким не відповідає обличчя, не піддаються модифікації.

Видалення треків. Для запобігання висковзування треків важливо вчасно їх видаляти. Для того щоб контролювати накопичення помилки, вводиться поняття накопичувального рівня довіри C_k .

Припустимо, що на k -м кадрі відомий рівень довіри $c_k \in (0, 1]$ (спосіб його оцінки обговорюється нижче). Тоді накопичувальний рівень довіри C_k визначається як добуток:

$$C_k = \prod_{g=0}^G c_{k-g}$$

Тут $G \geq 0$ – вік треку з моменту його останнього оновлення або створення.

Коли C_k пускається нижче заданого значення, відповідний трек видаляється. У момент створення або поновлення треку значення c_k і C_k встановлюються в 1.

При поступовому накопиченні помилки трекінгу значення C_k буде монотонно спадати і, в кінцевому рахунку, опуститься нижче порогового рівня. Таким чином, даний підхід запобігає висковзування треків. З іншого боку, в умовах сценарію відстеження безлічі облич трекі періодично оновлюються за допомогою детектора облич. Це запобігає помилкове видалення треків, що викликається змінами орієнтації і освітлення облич, які відслідковуються.

Міжкадрова підтримка треків. Оцінка стану обличчя на черговому кадрі є найважливішою складовою алгоритму трекінгу.

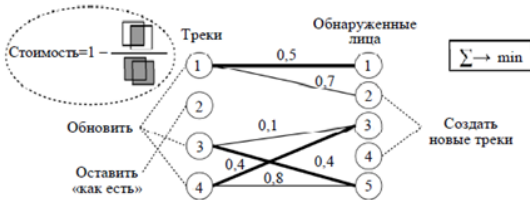


Рис. 2. Зіставлення виявлених облич та існуючих треків (приклад). Значення вартості (числа над ребрами) формуються на основі відносної площі перетину положень облич, зазначених детектором і трекером

4. Метод трекінгу на основі бінарних дескрипторів

У розділі розглядається метод, що дозволяє оцінити стан обличчя на поточному кадрі №(k), ґрунтуючись на відомому положенні цієї особи на попередньому кадрі №($k - 1$). Метод використовує бінарні дескриптори – локальні ознаки, представлені у формі бітових масивів. Вони обчислюються в ключових точках, які знаходяться за допомогою детектора ключових точок. В якості запобіжного відмінності бінарних дескрипторів використовується відстань Хеммінга. В роботі застосовується детектор ключових точок FAST (Features from Accelerated Segment

Test) і бінарні дескриптори BRIEF (Binary Robust Independent Elementary Features) [3].

Нижче приведена обчислювальна процедура, яка реалізує даний підхід:

- 1) В області обличчя на попередньому кадрі знаходяться ключові точки і обчислюються дескриптори;
- 2) Виходячи з припущень, введених в розділі «Постановка завдання», область обличчя з попереднього кадру розширюється. У цій розширеній області також знаходяться ключові точки і обчислюються дескриптори;
- 3) Виконується попарне порівняння дескрипторів з попереднього і поточного кадрів. Пари з відстанню Хеммінга менше заданого порогу T_H зберігаються і використовуються далі;
- 4) Обчислюється нове положення обличчя.

Розглянемо останній етап більш докладно. Після порівняння дескрипторів формується набір зіставлень: $(p_n, p'_n, w_n); n = 1 \dots N$. Тут $p_n = (x_n, y_n)$, $p'_n = (x'_n, y'_n)$ – координати ключових точок з попереднього і поточного кадрів відповідно. Ваги $w_n = [w_{min}, 1]$ обчислюються шляхом перетворення відповідних значень відстані Хеммінга, причому нульове відстань відповідає $w_n = 1$, а відстань T_H – в $w_n = [w_{min}]$.

Беручи до уваги введені вище припущення щодо розглянутого сценарію трекінгу, в модель руху включаються дві складові: здвиг $(\delta x, \delta y)$ і рівномірний масштаб s . Оцінка даних параметрів зводиться до наступної оптимізаційної задачі:

$$\theta = (\delta x, \delta y, s) = \operatorname{argmin} L_w(\theta)$$

$$L_\theta = \sum_{n=1}^N w_n \| f(p_n, \theta) - p'_n \|^2$$

$$f(p, \theta) = s \cdot p + (\delta x, \delta y)$$

Мінімальне значення цільової функції можна використовувати для формування оцінки рівня довіри c_k :

$$c_k(\theta) = \frac{1}{1 + \alpha \cdot \min L_w(\theta)} \in (0, 1]$$

Тут α – експериментально підбираємий параметр.

Так як модель руху володіє всього трьома ступенями свободи, обчислювальна процедура проста і стійка. З іншого боку, через використання простих ознак (BRIEF) може виникнути безліч некоректних зіставлень. Можливі шляхи вирішення даної проблеми включають досить низький (тобто жорсткий) поріг T_H , модифікацію ваг w_n (наприклад, з урахуванням відстані відповідних ключових точок від центру області обличчя), а також використання робастних методів оцінки параметрів, таких як, наприклад, RANSAC (RANdom SAmple Consensus).

5. Експериментальні дослідження

Метою експериментальних досліджень є оцінка швидкості роботи системи і якості трекінгу. Передбачається, що запропоноване рішення дозволяє значно збільшити швидкість обробки відеопотоку без істотного зниження якості трекінгу.

Представлена система реалізована на мові C++ з використанням бібліотек OpenCV і LEMON. Каскадний детектор з OpenCV навчений з параметрами, підібраними для високої швидкості обробки і достатньої якості виявлення.

Для оцінки характеристик розробленої системи використані дві відеопослідовності, що мають дозвіл 352×288 точок (рис. 3). Відео 1 є частиною бази SPEVI (рис. 2, а). Відео 2 являє собою чорно-білий запис з камери відеоспостереження (рис. 3, б). Для створення розмітки використана програма ViPER-GT.

Підвищення швидкості роботи, що досягається за рахунок використання трекера, залежить від кількості кадрів, що обробляються детектором (рис. 1). Чим більше кадрів доводиться на один запуск детектора (період запуску детектора, позначений символом τ_{det}), тим вище швидкість обробки кадрів всією системою. Експериментально отримані результати говорять про значне (чотириразовому) підвищенні швидкості роботи системи в порівнянні з базовою реалізацією.



Рис. 3. Використовувані для оцінки системи відеопослідовності: запис «motinas multi face frontal» з бази SPEVI (Відео 1) (а) і чорно-білий запис з камери відеоспостереження (Відео 2) (б)

Для оцінки якості трекінгу використані метрики Multiple Object Target Accuracy (MOTA) і Multiple Object Tracking Precision (MOTP) [4]. Метрика MOTA заснована на частоті пропусків облич (miss), частоті помилкових спрацьовувань (f_p) і частоті помилок призначення ідентифікатора (mme).

Відеозапис	Тестируемая система	miss	f_p , %	mme, %	MOTA	MOTP
Відео 1	Базова реалізація	8,40	0,57	0,19	0,91	0,29
	Система трекінга, $T_{det} = 3$	11,00	0,83	0,40	0,88	0,21
	Система трекінга, $T_{det} = 5$	13,00	0,35	0,28	0,86	0,20
Відео 2	Базова реалізація	29,00	1,40	0,46	0,70	0,25
	Система трекінга, $T_{det} = 3$	32,00	2,00	0,64	0,65	0,15
	Система трекінга, $T_{det} = 5$	36,00	1,50	0,40	0,62	0,15

Рис. 4. Результати оцінки якості трекінгу

Висновки

Запропонована підсистема трекінгу відрізняється простотою і гнучкістю. застосування ефективних алгоритмів обчислення бінарних дескрипторів і детектування ключових точок дозволяє обробляти відеопотік на високих швидкостях.

Проведені експериментальні дослідження показали, що запропонований підхід дозволяє домогтися значного (чотириразового) підвищення швидкості обробки в порівнянні з базовою реалізацією. При цьому якість трекінгу залишилося на допустимому рівні.

Розроблену систему можна успішно використовувати з різними детекторами облич (в тому числі досить повільними) для отримання повнофункціональної високошвидкісної системи трекінгу безлічі облич.

Завдяки підвищенню швидкості обробки алгоритм може знайти застосування не тільки в повномасштабних системах відеоспостереження, а й у вбудованих рішеннях, інтегрованих безпосередньо в камери відеоспостереження.

Перелік використаних джерел

1. Audiovisual liveness detection // Lecture Notes in Computer Science. / Melnikov A., Akhunzyanov R., Kudashev O., Luckyanets E. — 2015. — P. 643–652.
2. B. Minnehan, H. Spang, A.E. Savakis. robust and efficient tracker using dictionary of binary descriptors and locality constraints // Lecture Notes in Computer Science. — 2014. — P. 589–598.
3. Г.А. Кухарев, Е.И. Каменская, Ю.Н. Матвеев. Методы обработки и распознавания изображений лиц в задачах биометрии / под ред. М.В. Хитрова. — 2013. — С. 388.
4. K. Bernardin, R. Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics // EURASIP Journal on Image Video Processing. — 2008.